Local Convergence of the Proximal Point Method for a Special Class of Nonconvex Functions on Hadamard Manifolds

G. C. Bento* O. P. Ferreira † P. R. Oliveira ‡

April 10, 2010

Abstract

Local convergence analysis of the proximal point method for special class of nonconvex function on Hadamard manifold is presented in this paper. The well definedness of the sequence generated by the proximal point method is guaranteed. Moreover, is proved that each cluster point of this sequence satisfies the necessary optimality conditions and, under additional assumptions, its convergence for a minimizer is obtained.

Key words: proximal point method, nonconvex functions, Hadamard manifolds.

1 Introduction

The extension of the concepts and techniques of the Mathematical Programming of the Euclidean space R^n to Riemannian manifolds is natural. It has been frequently done in recent years, with a theoretical purpose and also to obtain effective algorithms; see [1], [3], [10], [11], [16], [19], [20], [23] and [25]. In particular, we observe that, these extensions allow the solving some nonconvex constrained problems in Euclidean space. More precisely, nonconvex problems in the classic sense may become convex with the introduction of an adequate Riemannian metric on the manifold (see, for example [9]). The proximal point algorithm, introduced by Martinet [17] and Rockafellar [21], has been extended to different contexts, see [11], [19] and their references. In [11] the authors generalized the proximal point method for solve convex optimization problems of the form

$$(P) \quad \begin{array}{l} \min f(p) \\ \text{s.t. } p \in M, \end{array} \tag{1}$$

*IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, BR (Email: glaydston@mat.ufg.br)

†IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, BR (Email: orizon@mat.ufg.br). The author was supported in part by CNPq Grant 302618/2005-8, PRONEX–Optimization(FAPERJ/CNPq) and FUNAPE/UFG.

‡COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, BR (Email: poliveir@cos.ufrj.br). This author was supported in part by CNPq.

where M is a Hadamard manifold and $f : M \to \mathbb{R}$ is a convex function (in the Riemannian sense). The method was described as follows:

$$p^{k+1} := \text{argmin}_{p\in M} \left\{ f(p) + \frac{\lambda_k}{2} d^2(p, p^k) \right\}, \qquad (2)$$ E:1.2

with $p^\circ \in M$ an arbitrary point, d the intrinsic Riemannian distance (to be defined later on) and $\{\lambda_k\}$ a sequence of positive numbers. The authors also showed that this extension is natural. As regards to [19] the authors generalized the proximal point method with Bregman distance to solve quasiconvex and convex optimization problems also on Hadamard manifold. Spingarn in [24] has, in particular, developed the proximal point method for the minimization of a certain class of nondifferentiable noncovex functions, namely, the lower-C^2 functions defined in Euclidean spaces, see also [13]. Kaplan and Tichatschke in [15] also applied the proximal point method for the minimization of a similar class of the ones of [13] and [24], namely, functions defined as maximum of a certain collection (finite/infinite) of continuously differentiable functions. In [2] we study, in the Riemannian context, the same class of functions studied in [15]. In that context we applied the proximal point method (2) to solve the problem (1), however we assumed that the collection of functions defining the objective function was finite.

Our goal is to extend the results of [2]. We consider that the objective function is given by the maximum of a collection infinite of continuously differentiable functions. To obtain the results in [2], it was necessary to study the generalized directional derivative in the Riemannian manifolds context. In this paper we go further in the study of properties of the generalized directional derivative in order to analyze the convergence of the proximal point method. Several works have studied such concepts and presented many useful results in the Riemannian optimization context, see for example [4], [16], [18] and [26].

The paper is divided as follows. In Section 1.1 we give the notation and some results on the Riemannian geometry which we will use along the paper. In Section 2 we recall some facts of the convex analysis on Hadamard manifolds. In Section 3 we present definition of generalized directional derivative of a locally Lipschitz function (not necessarily convex) which, in the Euclidean case, coincides with the Clarke's generalized directional derivative. Moreover, some properties of that derivative are presented, amongst which the upper semicontinuity of the directional derivative. In Section 4 we study the proximal point method (2) to solve the problem (1), in the case where the objective function is a real-valued function (non necessarily convex) on a Hadamard manifold M given by the maximum of a certain class of functions. Finally in Section 5 we provide an example where the proximal point method for nonconvex problems is applied.

sec2

1.1 Notation and terminology

In this section we introduce some fundamental properties and notations on Riemannian geometry. These basics facts can be found in any introductory book on Riemannian geometry, such as in [6] and [22].

Let M be a n-dimentional connected manifold. We denote by T_pM the n-dimentional *tangent space* of M at p, by $TM = \cup_{p\in M} T_pM$ *tangent bundle* of M and by $\mathcal{X}(M)$ the space of smooth vector fields over M. When M is endowed with a Riemannian metric $\langle\, ,\, \rangle$, with the corresponding norm denoted by $\|\;\|$, then M is now a Riemannian manifold. Recall that the metric can be used to define the lenght of piecewise smooth curves $\gamma : [a, b] \to M$ joining p to q, i.e., such that $\gamma(a) = p$ and $\gamma(b) = q$, by

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt,$$

and, moreover, by minimizing this length functional over the set of all such curves, we obtain a Riemannian distance $d(p, q)$ which induces the original topology on M. The metric induces a map $f \mapsto \operatorname{grad} f \in \mathcal{X}(M)$ which associates to each smooth function on M its gradient via the rule $\langle \operatorname{grad} f, X\rangle = df(X),\ X \in \mathcal{X}(M)$. Let ∇ be the Levi-Civita connection associated to $(M, \langle\, ,\, \rangle)$. A vector field V along γ is said to be *parallel* if $\nabla_{\gamma'} V = 0$. If γ' itself is parallel we say that γ is a *geodesic*. Given that geodesic equation $\nabla_{\ \gamma'} \gamma' = 0$ is a second order nonlinear ordinary differential equation, then geodesic $\gamma = \gamma_v(., p)$ is determined by its position p and velocity v at p. It is easy to check that $\|\gamma'\|$ is constant. We say that γ is *normalized* if $\|\gamma'\| = 1$. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. A geodesic segment joining p to q in M is said to be *minimal* if its length equals $d(p, q)$ and this geodesic is called a *minimizing geodesic*. If γ is a curve joining points p and q in M then, for each $t \in [a, b]$, ∇ induces a linear isometry, relative to $\langle\, ,\, \rangle$, $P_{\gamma(a)\gamma(t)} : T_{\gamma(a)}M \to T_{\gamma(t)}M$, the so-called *parallel transport* along γ from $\gamma(a)$ to $\gamma(t)$. The inverse map of $P_{\gamma(a)\gamma(t)}$ is denoted by $P^{-1}_{\gamma(a)\gamma(t)} : T_{\gamma(t)}M \to T_{\gamma(a)}M$. In the particular case of γ is the unique curve joining points p and q in M then parallel transport along γ from p to q is denoted by $P_{pq} : T_pM \to T_qM$.

A Riemannian manifold is *complete* if geodesics are defined for any values of t. Hopf-Rinow's theorem asserts that if this is the case then any pair of points, say p and q, in M can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, (M, d) is a complete metric space and bounded and closed subsets are compact. Take $p \in M$. The *exponential map* $exp_p : T_pM \to M$ is defined by $exp_p v = \gamma_v(1, p)$.

We denote by R *the curvature tensor* defined by $R(X, Y) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[Y,X]} Z$, with $X, Y, Z \in \mathcal{X}(M)$, where $[X, Y] = YX - XY$. Then the *sectional curvature* with respect to X and Y is given by $K(X, Y) = \langle R(X, Y)Y, X\rangle/(||X||^2||X||^2 - \langle X\, ,\, Y\rangle^2)$, where $||X|| = \langle X, X\rangle^{1/2}$. If $K(X, Y) \leqslant 0$ for all X and Y, then M is called a *Riemannian manifold of nonpositive curvature* and we use the short notation $K \leqslant 0$.

T:Hadamard **Theorem 1.1.** *Let M be a complete, simply connected Riemannian manifold with nonpositive sectional curvature. Then M is diffeomorphic to the Euclidean space $\mathbb{R}^n$, $n = dim M$. More precisely, at any point $p \in M$, the exponential mapping $exp_p : T_pM \to M$ is a diffeomorphism.*

Proof. See [6] and [22]. ☐

A complete simply connected Riemannian manifold of nonpositive sectional curvature is called a *Hadamard manifold*. The Theorem 1.1 says that if M is Hadamard manifold, then M has the same topology and differential structure of the Euclidean space $\mathbb{R}^n$. Furthermore, are known some similar geometrical properties of the Euclidean space $\mathbb{R}^n$, such as, given two points there exists an unique geodesic that joins them. *In this paper, all manifolds M are assumed to be Hadamard finite dimensional.*

2 Convexity in Hadamard manifold

sec3

In this section, we introduce some fundamental properties and notations of convex analysis on Hadamard manifolds which will be used later. We will see that these properties are similar to those obtained in convex analysis on the Euclidean space $\mathbb{R}^n$. References to convex analysis on Euclidean space $\mathbb{R}^n$ are in [14], and on Riemannian manifold are in [7], [11], [20], [22], [23] and [25].

The set $\Omega \subset M$ is said to be *convex* if any geodesic segment with end points in Ω is contained in Ω. Let $\Omega \subset M$ be an open convex set. A function $f : M \to \mathbb{R}$ is said to be *convex* (respectively, *strictly convex*) on Ω if for any geodesic segment $\gamma : [a, b] \to \Omega$ the composition $f \circ \gamma : [a, b] \to \mathbb{R}$ is convex (respectively, strictly convex). Now, a function $f : M \to \mathbb{R}$ is said to be *strongly convex* on Ω with constant $L > 0$ if, for any geodesic segment $\gamma : [a, b] \to \Omega$, the composition $f \circ \gamma : [a, b] \to \mathbb{R}$ is strongly convex with constant $L\|\gamma'(0)\|^2$. Take $p \in M$. A vector $s \in T_pM$ is said to be a *subgradient* of f at p if

$$f(q) \geq f(p) + \langle s, \exp_p^{-1} q\rangle,$$

for any $q \in M$. The set of all subgradients of f at p, denoted by $\partial f(p)$, is called the *subdifferential* of f at p.

Take $p \in M$. Let $exp_p^{-1} : M \to T_pM$ be the inverse of the exponential map which is also C^∞. Note that $d(q\,,\,p) = ||exp_p^{-1}q||$, the map $d^2(\,.\,, p)\colon M \to \mathbb{R}$ is C^∞ and

$$\operatorname{grad} \frac{1}{2} d^2(q, p) = -exp_q^{-1}p,$$

(remember that M is a Hadamard manifold) see, for example, [22].

FunDistConv **Proposition 2.1.** *Take $p \in M$. The map $d^2(\,.\,, p)/2$ is strongly convex.*

Proof. See [7]. □

def2.14 **Definition 2.1.** *Let $\Omega \subset M$ be an open convex set. A function $f : M \to \mathbb{R}$ is said to be Lipschitz on Ω if there exists a constant $L := L(\Omega) \geq 0$ such that*

$$|f(p) - f(q)| \leq Ld(p, q), \qquad p, q \in \Omega. \tag{3}$$

Lipsch1

Moreover, if it is established that for all $p_0 \in \Omega$ there exists $L(p_0) \geq 0$ and $\delta = \delta(p_0) > 0$ such that the inequality (3) occurs with $L = L(p_0)$ for all $p, q \in B_\delta(p_0) := \{p \in \Omega : d(p, p_0) < \delta\}$, then f is called locally Lipschitz on Ω.

obs2.15 **Remark 2.1.** *As an immediate consequence of the triangular inequality we obtain that* $|d(p,p_0) - d(q,p_0)| \leq d(p,q)$ *for all* p, q *and* $p_0 \in M$. *Then, of the Definition 2.1, we get that the Riemannian distance function to a fixed point,* $d(\cdot, q)$ *is Lipschitzian and therefore Lipschitzian locally. In fact, it well known that every convex function is locally Lipschitz and consequently continuous. See [12].*

SubClarke2 **Proposition 2.2.** *Let* $\Omega \subset M$ *be a open convex set,* $f : M \to \mathbb{R}$ *and* $p \in M$. *If there exists* $\lambda > 0$ *such that* $f + (\lambda/2)\, d^2(.\,,p) : M \to \mathbb{R}$ *is convex on* Ω, *then* f *is Lipschitz locally on* Ω.

Proof. Because $f + (\lambda/2)\, d^2(.\,,p)$ is convex, it follows from Remark 2.1 that for any $\tilde{p} \in \Omega$ there exist $L_1, \delta_1 > 0$ such that

$$\left|[f(q_1) + (\lambda/2)\, d^2(q_1\,,p)] - [f(q_2) + (\lambda/2)\, d^2(q_2\,,p)]\right| \leq L_1 d(q_1,q_2), \qquad \forall\, q_1,q_2 \in B(\tilde{p},\delta_1). \tag{4}$$ DesLip100

Moreover, Proposition 2.1 together with Remark 2.1 imply that there exist $L_2, \delta_2 > 0$ such that

$$|(1/2)d^2(q_1\,,p) - (1/2)d^2(q_2\,,p)| \leq L_2 d(q_1,q_2), \qquad \forall\, q_1,q_2 \in B(\tilde{p},\delta_1). \tag{5}$$ DesLip101

Simples algebraic manipulations implies that

$$\begin{aligned}|f(q_1) - f(q_2)| \leq \left|[f(q_1) + \lambda/2)\, d^2(q_1\,,p)] - [f(q_2) + (\lambda/2)\, d^2(q_2\,,p)]\right| + \\ + \left|(\lambda/2)\, d^2(q_2\,,p) - (\lambda/2)\, d^2(q_1\,,p)\right|.\end{aligned}$$

Therefore, taking $\delta = \min\{\delta_1, \delta_2\}$, using (4) and (5) we conclude from last inequality that

$$|f(q_1) - f(q_2)| \leq (L_1 + \lambda L_2)d(q_1,q_2), \qquad \forall\, q_1,q_2 \in B(\tilde{p},\delta),$$

and the proof is finished. □

Definition 2.2. *Let* $\Omega \subset M$ *be a open convex set and* $f : M \to \mathbb{R}$ *a continuously differentiable function on* Ω. *The gradient vector field* grad f *is said to be Lipschitz with constant* $\Gamma \geq 0$ *on* Ω *always that*

$$\|\operatorname{grad} f(q) - P_{pq} \operatorname{grad} f(p)\| \leq \Gamma d(p,q), \qquad p,q \in \Omega,$$

where P_{pq} *is the parallel transport along the geodesic segment joining* p *to* q.

3 Generalized directional derivatives

sec:dd

In this section we present definitions for the generalized directional derivative and subdifferential of a locally Lipschitz function (not necessarily convex) which, in the Euclidean case, coincide with the Clarke's generalized directional derivative and subdifferential, respectively. Moreover, some properties of those concepts are presented, amongst them the upper semicontinuity of the directional derivative and a relationship between the subdifferential of a sum of two Lipschitz locally function (in the particular case that one of them is differentiable) and its subdifferentials.

d:Clarke **Definition 3.1.** *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a locally Lipschitz function on Ω. The generalized directional derivative $f^\circ : T\Omega \to \mathbb{R}$ of f is defined by*

$$f^\circ(p, v) := \limsup_{t \downarrow 0 \;\; q \to p} \frac{f\left(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v\right) - f(q)}{t}, \tag{6}$$ Clarke1

where $(D\exp_p)_{\exp_p^{-1} q}$ denotes the differential of $\exp_p$ at $\exp_p^{-1} q$.

It is worth to pointed out that an equivalently definition has appeared in [4].

Remark 3.1. *The generalized directional derivative is well defined. Indeed, let $L_p > 0$ the Lipschitz constant of f at p and $\delta = \delta(p) > 0$ such that*

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \leq L_p\, d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v,\, q), \quad q \in B_\delta(p), \quad t \in [0, \delta).$$

Since $d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v,\, q) = t\|(D\exp_p)_{\exp_p^{-1} q} v\|$, above inequality becomes

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \leq L_p\, t\|(D\exp_p)_{\exp_p^{-1} q} v\|, \quad q \in B_\delta(p), \quad t \in [0, \delta).$$

Since $\lim_{q \to p} (D\exp_p)_{\exp_p^{-1} q} v = v$ our statement follows from last inequality.

Remark 3.2. *Note that, if $M = \mathbb{R}^n$ then $\exp_p w = p + w$ and $(D\exp_p)_{\exp_p^{-1} q} v = v$. In this case,* (6) *becomes*

$$f^\circ(p, v) = \limsup_{t \downarrow 0 \;\; q \to p} \frac{f(q + tv) - f(q)}{t},$$

which is the Clarke's generalized directional derivative, see [5]. Therefore, the generalized differential derivative on Hadamard manifold is a natural extension of the Clarke's generalized differential derivative.

Now we are going to prove the upper semicontinuity of the generalized directional derivative.

PropoConverg **Proposition 3.1.** *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ be a locally Lipschitz function. Then, f° is upper semicontinuous on $T\Omega$, i.e., if $(p, v) \in T\Omega$ and $\{p^k, v^k\}$ is a sequence in $T\Omega$ such that $\lim_{k \to +\infty}(p^k, v^k) = (p, v)$, then*

$$\limsup_{k \to +\infty} f^\circ(p^k, v^k) \leq f^\circ(p, v). \tag{7}$$ DerFinita2

Proof. Let $(p, v) \in T\Omega$ and $\{(p^k, v^k)\} \subset T\Omega$ such that $\lim_{k \to +\infty}(p^k, v^k) = (p, v)$. For proving the inequality (7) first note that for each k

$$f^\circ(p^k, v^k) \leq \limsup_{t \downarrow 0 \;\; (q,w) \to (p^k, v^k)} \frac{f(\exp_q tw) - f(q)}{t}, \qquad (q, w) \in T\Omega.$$

So, by definition of upper limit, there exists $(q^k, w^k) \in T\Omega - \{(p^k, v^k)\}$ and $t_k > 0$ such that

$$f^\circ(p^k, v^k) - \frac{1}{k} < \frac{f(\exp_{q^k} t_k w^k) - f(q^k)}{t_k}, \qquad \tilde{d}((q^k, w^k), (p^k, v^k)) + t_k < \frac{1}{k}, \tag{8}$$ UpperSem1

with $\tilde{d}$ being the Riemannian distance in TM. Let $U_p \subset \Omega$ be a neighborhood of p such that $TU_p \approx U_p \times \mathbb{R}^n$, f is Lipschitz in U_p with constant L_p and exp is Lipschitz on TU_p with constant K. From the first inequality in (8), we obtain

$$f^\circ(p^k, v^k) - \frac{1}{k} < \frac{f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) - f(q^k)}{t_k} + \frac{f(\exp_{q^k} t_k w^k) - f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right)}{t_k}. \tag{9}$$ UpperSem3

On the other hand, as $\lim_{k\to+\infty}(p^k, v^k) = (p, v)$, we conclude from the second inequality in (8) that

$$\exp_{q^k} t_k w^k \in U_p, \qquad \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v) \in U_p, \qquad k > k_0,$$

for k_0 sufficiently large. Thus, as f is Lipschitz on U_p, for $k > k_0$ we have

$$\left| f(\exp_{q^k} t_k w^k) - f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) \right| \leq L_p\, d\left(\exp_{q^k} t_k w^k,\, \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right). \tag{10}$$ UpperSemI

Now, taking into account that exp is Lipschitz on TU_p, in the particular case that $k > k_0$

$$d\left(\exp_{q^k} t_k w^k,\, \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) \leq k \| t_k w^k - t_k (D\exp_p)_{\exp_p^{-1} q^k} v \|.$$

Since $\lim_{k\to+\infty} p^k = p$, second equation in (8) imply that $\lim_{k\to+\infty} q^k = p$. Consequently,

$$\lim_{k\to+\infty} (D\exp_p)_{\exp_p^{-1} q_k} v = v,$$

which, together with last inequality imply

$$\lim_{k\to+\infty} d\left(\exp_{q^k} t_k w^k,\, \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) / t_k = 0.$$

Therefore, combining last equation, (9), (10), and Definition 3.1 the result follows. $\square$

Next we generalize the definition of subdifferential for locally Lipschitz functions defined on Hadamard manifold.

Definition 3.2. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a locally Lipschitz function on Ω. The generalized subdifferential of f at $p \in \Omega$, denoted by $\partial^\circ f(p)$, is defined by*

$$\partial^\circ f(p) := \big\{ w \in T_pM : f^\circ(p, v) \geq \langle w, v\rangle, \forall\, v \in T_pM \big\}.$$

regular 1 **Remark 3.3.** *Let $\Omega \subset M$ be an open convex set. If the function $f : M \to \mathbb{R}$ is convex on Ω, then $f^\circ(p, v) = f'(p, v)$ (respectively, $\partial^\circ f(p) = \partial f(p)$) for all $p \in \Omega$, i.e., the directional derivatives (respectively, subdifferential) for Lipschitz functions is a generalization of the directional derivatives (respectively, subdifferential) for convex functions. See [4] Claim 5.4 in the proof of Theorem 5.3.*

Definition 3.3. *Let $f : M \to \mathbb{R}$ be locally Lipschitz function. A point $p \in \Omega$ is said to be a stationary point of f always $0 \in \partial^\circ f(p)$.*

lemacl1 **Lemma 3.1.** *Let $\Omega \subset M$ be an open set. If $f : M \to \mathbb{R}$ is locally Lipschitz function on Ω and $g : M \to \mathbb{R}$ is convex on Ω, then*

$$(f + g)^\circ(p, v) = f^\circ(p, v) + g'(p, v) \qquad p \in \Omega, \quad v \in T_pM. \tag{11}$$ ddg:1

Moreover, if g is differentiable, we have

$$\partial^\circ(f + g)(p) = \partial^\circ f(p) + \operatorname{grad} g(p), \qquad p \in \Omega. \tag{12}$$ sdg:1

Proof. Using the definition of the generalized directional derivative and simple algebraic manipulations, we obtain

$$(f+g)^\circ(p,v) = \limsup_{t\downarrow 0\ q\to p} \left[\frac{f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)}{t} + \frac{g(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - g(q)}{t} \right].$$

From basic properties of the upper limit along with the definition of directional derivative generalized and Remark 3.3, follows that

$$(f + g)^\circ(p, v) \leq f^\circ(p, v) + g'(p, v). \tag{13}$$ P:ddg1

Now, as $f^\circ(p, v) = \left((f + g) + (-g)\right)^\circ (p, v)$, above inequality implies, in particular

$$f^\circ(p, v) \leq (f + g)^\circ(p, v) + (-g)'(p, v),$$

which is equivalent to

$$(f + g)^\circ(p, v) \geq f^\circ(p, v) + g'(p, v).$$

Thus, combining last inequality with inequality (13), the equality (11) is obtained.

In the case that g is also differentiable, in particular $g'(p, v) = \langle \operatorname{grad} g(p), v\rangle$. Therefore, the proof of the equality (12) is an immediate consequence of the equality (11) along with the definition of the generalized subdifferential. $\square$

c:ssfd **Corollary 3.1.** *Let $\Omega \subset M$ be a open convex set, $f : M \to \mathbb{R}$ be locally Lipschitz functions on Ω, $\tilde{p} \in M$ and $\lambda > 0$ such that $f + (\lambda/2)\, d^2(.\,, \tilde{p}) : M \to \mathbb{R}$ is convex on Ω. If $p \in \Omega$ is a minimizer of $f + (\lambda/2)d^2(., \tilde{p})$ then*

$$\lambda \exp_p^{-1} \tilde{p} \in \partial^\circ f(p).$$

Proof. Since p is a minimizer of $f + (\lambda/2)d^2(., \tilde{p})$ we obtain

$$0 \in \partial \left(f + \frac{\lambda}{2} d^2(.\,, \tilde{p}) \right)(p). \tag{14}$$ eq:mc1

On the other hand, as $f + (\lambda/2)\, d^2(.\,, \tilde{p})$ is convex on Ω and $(\lambda/2)\, d^2(.\,, \tilde{p})$ is differentiable with $\operatorname{grad}(\lambda/2)\, d^2(q, p) = -\lambda \exp_q^{-1} p$, using Remark 3.3, Proposition 2.2 and applying Lema 3.1 with $g = (\lambda/2)\, d^2(.\,, \tilde{p})$, we have

$$\partial \left(f + \frac{\lambda}{2} d^2(., \tilde{p}) \right)(p) = \partial^\circ \left(f + \frac{\lambda}{2} d^2(., \tilde{p}) \right)(p) = \partial^\circ f(p) - \lambda \exp_p^{-1} \tilde{p}. \tag{15}$$ eq:mc2

Therefore, the result follows by combining (14) with (15). □

4 Proximal Point Method for Nonconvex Problems

sec5

In this section we present an application of the proximal point method for minimize a real-valued function (non necessarily convex) given by the maximum of a certain class of continuously differentiable functions. Our goal is to prove the following theorem:

MPP10 **Theorem 4.1.** *Let $\Omega \subset M$ be an open convex set, $q \in M$ and $T \subset \mathbb{R}$ a compact set. Let $\varphi : M \times T \to \mathbb{R}$ be a continuous function on $\Omega \times T$ such that $\varphi(., \tau) : M \to \mathbb{R}$ is a continuously differentiable function on $\bar{\Omega}$ (closure of Ω) for all $\tau \in T$, and $f : M \to \mathbb{R}$ defined by*

$$f(p) := \max_{\tau \in T} \varphi(p, \tau).$$

Assume that $-\infty < \inf_{p \in M} f(p)$, $\operatorname{grad}_p \varphi(., \tau)$ is Lipschitz on Ω with constant L_τ for each $\tau \in T$ such that $\sup_{\tau \in T} L_\tau < +\infty$ and

$$L_f(f(q)) = \{p \in M : f(p) \le f(q)\} \subset \Omega, \qquad \inf_{p \in M} f(p) < f(q).$$

Take $0 < \bar{\lambda}$ and a sequence $\{\lambda_k\}$ satisfying $\sup_{\tau \in T} L_\tau < \lambda_k \le \bar{\lambda}$ and $\hat{p} \in L_f(f(q))$. Then the proximal point method

$$p^{k+1} := \operatorname{argmin}_{p \in M} \left\{ f(p) + \frac{\lambda_k}{2} d^2(p, p^k) \right\}, \qquad k = 0, 1, \ldots, \tag{16}$$ E:1.22

with starting point $p^0 = \hat{p}$ is well defined, the generated sequence $\{p^k\}$ rests in $L_f(f(q))$ and satisfies only one of the following statement

i) $\{p^k\}$ *is finite, i.e.,* $p^{k+1} = p^k$ *for some* k *and, in this case,* p^k *is a stationary point of* f*,*

ii) $\{p^k\}$ *is infinite and, in this case, any cluster point of* $\{p^k\}$ *is a stationary point of* f*.*

Moreover, assume that the minimizer set of f *is non-empty, i. e.,*

h1) $U^* = \{p : f(p) = \inf_{p\in M} f(p)\} \neq \emptyset$*.*

Let $c \in (\inf_{p\in M} f(p),\ f(q))$*. If, in addition, the following assumptions hold:*

h2) $L_f(c)$ *is convex,* f *is convex on* $L_f(c)$ *;*

h3) *for all* $p \in L_f(f(q)) \setminus L_f(c)$ *and* $y(p) \in \partial^\circ f(p)$ *we have* $\|y(p)\| > \delta > 0$*,*

then the sequence $\{p^k\}$ *generated by* (16) *with*

$$\sup_{\tau\in T} L_\tau < \lambda_k \leq \bar{\lambda}, \qquad k = 0, 1, \ldots \tag{17}$$ mpp102

converge to a point $p^* \in U^*$*.*

Remark 4.1. *The continuity of each function* $\varphi(.,\tau)$ *on* $\bar{\Omega}$ *in* **h2** *guarantees that the level sets of the fuction f, in particular the solution set* U^**, are closed in the topology of the manifold* M*.*

In the next remark we show that if Ω is bounded and $\varphi(.,\tau)$ is convex on Ω for all $\tau \in T$ then f satisfies the assumptions **h2** and **h3**.

re:tconver **Remark 4.2.** *If* $\varphi(.,\tau)$ *is a convex function on* Ω *for all* $\tau \in T$ *then the assumtion* **h2** *is naturally verified and if* **h1** *hold then* **h3** *also holds. For details, see [2].*

In order to prove above theorem we need of some preliminary results. From now on we assume that every assumptions on Theorem 4.1 hold, with the exception of **h1**, **h2** and **h3**, which will be considered to hold only when explicitly stated.

MPP8 **Lemma 4.1.** *For all* $\tilde{p} \in M$ *and* λ *satisfying*

$$\sup_{\tau\in T} L_\tau < \lambda,$$

function $f + (\lambda/2)d^2(.,\tilde{p})$ *is strongly convex on* Ω *with constant* $\lambda - \sup_{\tau\in T} L_\tau$*.*

Proof. Since T is compact and φ is continuous the well definition of f follows. To conclude, see Lemma 4.1 in [2]. $\square$

cor:wdf **Corollary 4.1.** *The proximal point method* (16) *applied to* f *with starting point* $p^0 = \hat{p}$ *is well defined.*

Proof. Since compactness play no rule, the proof is equal to the proof of Corollary 4.1 in [2]. ◻

mpprox10 **Lemma 4.2.** *Let $\{p^k\}$ be the sequence generated by the proximal point method* (16)*. Then*

i) $0 \in \partial\left(f + \frac{\lambda_k}{2} d^2(.\,,\, p^k)\right)(p^{k+1}), \quad k = 0, 1, \ldots.$

ii) $\lim\limits_{s\to\infty} d(p^{k+1}, p^k) = 0.$

Moreover, if λ_k satisfies (17) *and* **h1**, **h2** *and* **h3** *hold, then $\{p^k\}$ converges to a point $p^* \in U^*$.*

Proof. Since compactness play no rule, the proof is similar to the proof of Lemmas 4.2, 4.3 and 4.4 of [2]. ◻

Proof of Theorem 4.1

Proof. The well definition of the proximal point method (16) follows from the Corollary 4.1. Let $\{p^k\}$ be the sequence generated by proximal point method. Because $p^0 = \hat{p} \in L_f(f(q))$, (16) implies that the whole sequence is in $L_f(f(q))$. From item i of Lemma 4.2, we have

$$0 \in \partial\left(f + \frac{\lambda_k}{2} d^2(.\,,\, p^k)\right)(p^{k+1}), \quad k = 0, 1, \ldots.$$

Since $\sup_{\tau\in T} L_\tau < \lambda_k$, Lemma 4.1 implies that $f + (\lambda_k/2)d^2(.\,, p^k)$ is strongly convex on Ω, which together with Proposition 2.2 give us that f is locally Lipschitz on Ω. So, using the definition of p^{k+1}, we conclude from Corollary 3.1 with $\lambda = \lambda_k$, $\tilde{p} = p^k$ and $p = p^{k+1}$ that

$$\lambda_k \exp^{-1}_{p^{k+1}} p^k \in \partial^\circ f(p^{k+1}). \tag{18}$$
MPP12

If $\{p^k\}$ is finite, then $p^{k+1} = p^k$ for some k and latter inclusion implies that $0 \in \partial^\circ f(p^{k+1})$, i.e., p^k is a stationary point of f. Now assume that $\{p^k\}$ is a infinite sequence. If $\bar{p}$ is a cluster point of $\{p^k\}$, then there exists a subsequence $\{p^{k_s}\}$ of $\{p^k\}$ such that $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ and item ii of Lemma 4.2 implies

$$\lim_{s\to\infty} \|\exp^{-1}_{p^{k_s+1}} p^{k_s}\| = \lim_{s\to\infty} d(p^{k_s+1}, p^{k_s}) = 0. \tag{19}$$
AcumEstac1

Now, from the relation (18), we have

$$f^\circ(p^{k_s+1}, v) \geq \lambda_{k_s} \langle \exp^{-1}_{p^{k_s+1}} p^{k_s}, v\rangle, \qquad \forall\, v \in T_{p^{k_s+1}}M.$$

Let $\bar{v} \in T_{\bar{p}}M$. Hence, latter inequality implies that

$$f^\circ(p^{k_s+1}, v^{k_s+1}) \geq \lambda_{k_s} \langle \exp^{-1}_{p^{k_s+1}} p^{k_s}, v^{k_s+1}\rangle, \qquad v^{k_s+1} = D(\exp_{\bar{p}})_{\exp^{-1}_{\bar{p}} p^{k_s+1}} \bar{v}.$$

Note that $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ implies $\lim_{s\to+\infty} v^{k_s+1} = \bar{v}$. Because $\{\lambda_{k_s}\}$ is bounded, letting s goes to $+\infty$ in the last inequality, Proposition 3.1 together with (19) give us

$$f^\circ(\bar{p}, \bar{v}) \geq \lim_{s\to+\infty} \sup f^\circ(p^{k_s+1}, v^{k_s+1}) \geq 0,$$

which implies that $0 \in \partial^\circ f(\bar{p})$, i.e., $\bar{p}$ is a stationary point of f and the first part of the theorem is concluded.

The second part follows from the last part of Lemma 4.2 and the proof of the theorem is finished. $\square$

5 Example

sec6

Let $(\mathbb{R}_{++}, \langle\, , \,\rangle)$ be the Riemannian manifold, where $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$ and $\langle\, , \,\rangle$ is the Riemannian metric $\langle u, v\rangle = g(x)uv$ with $g : \mathbb{R}_{++} \to (0, +\infty)$. So, the Christoffel symbol and the geodesic equation are given by

$$\Gamma(x) = \frac{1}{2} g^{-1}(x) \frac{dg(x)}{dx} = \frac{d}{dx} \ln \sqrt{g(x)}, \qquad \frac{d^2x}{dt^2} + \Gamma(x) \left(\frac{dx}{dt}\right)^2 = 0,$$

respectively. Besides, in relation to the twice differentiable function $h : \mathbb{R}_{++} \to \mathbb{R}$, the Gradient and the Hessian of h are given by

$$\operatorname{grad} h = g^{-1} h', \qquad \text{hess } h = h'' - \Gamma h',$$

respectively, where h' and h'' denote the first and second derivatives of h in the Euclidean sense. For more details see [25]. So, in the particular case of $g(x) = x^{-2}$,

$$\Gamma(x) = -x^{-1}, \quad \operatorname{grad} h(x) = x^2 h'(x), \quad \text{hess } h(x) = h''(x) + x^{-1} h'(x). \tag{20}$$ Hess:1

Moreover, the map $\psi : \mathbb{R} \to \mathbb{R}_{++}$ defined by $\psi(x) = \mathrm{e}^x$ is an isometry between the Euclidean space $\mathbb{R}$ and the manifold $(\mathbb{R}_{++}, \langle\, , \,\rangle)$, and the Riemannian distance $d : \mathbb{R}_{++} \times \mathbb{R}_{++} \to \mathbb{R}_{+}$ is given by

$$d(x, y) = |\psi^{-1}(x) - \psi^{-1}(y)| = |\ln(x/y)|, \tag{21}$$ dRiem:1

see, for example [9]. Therefore, $(\mathbb{R}_{++}, \langle\, , \,\rangle)$ is a Hadamard manifold and the unique geodesic $x : \mathbb{R} \to \mathbb{R}_{++}$ with initial conditions $x(0) = x_0$ and $x'(0) = v$ is given by

$$x(t) = x_0 \mathrm{e}^{(v/x_0)t}.$$

Now let $f_1, f_2, f : \mathbb{R}_{++} \to \mathbb{R}$ and $\varphi : \mathbb{R}_{++} \times [0, 1] \to \mathbb{R}$ be real-valued functions such that

$$\varphi(x, \tau) = f_1(x) + t(f_2(x) - f_1(x)), \quad f(x) = \max_{\tau\in[0,1]} \varphi(x, \tau), \tag{22}$$ eq:dfs

and consider the problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in \mathbb{R}_{++}. \end{aligned}$$

Take a sequence $\{\lambda_k\}$ satisfying $0 < \lambda_k$. From (21), the proximal point method (16) becomes

$$x^{k+1} := \operatorname{argmin}_{x \in \mathbb{R}_{++}} \left\{ f(x) + \frac{\lambda_k}{2} \ln^2 \left(\frac{x}{x^k} \right) \right\}, \qquad k = 0, 1, \ldots .$$

If f_1 and f_2 are given, respectively, by $f_1(x) = \ln(x)$ and $f_2(x) = -\ln(x) + \mathrm{e}^{-2x} - \mathrm{e}^{-2}$, then φ is continuous and $\varphi(., \tau)$ is continuously differentiable for each $\tau \in [0, 1]$. The last expression in (20) implies that

$$\operatorname{hess} f_1(x) = 0 \quad , \quad \operatorname{hess} f_2(x) = (4 - 2/x)\mathrm{e}^{-2x}, \qquad x \in \mathbb{R}_{++}, \tag{23}$$

ExHess:1

e, as a consequence, first expression in (22) give us

$$\operatorname{hess}_x \varphi(x, \tau) = \tau \operatorname{hess} f_2(x), \quad \forall\, x \in \mathbb{R}_{++} \quad \forall\, \tau \in [0, 1].$$

Note that, for $0 < \epsilon < 1/4$ and $\Omega = (\epsilon, +\infty)$, hess f_2 is bounded on Ω and therefore grad f_2 is Lipschitz on Ω. We denote by L the constant of Lipschitz of grad f_2. From the last equality $\operatorname{hess}_x \varphi(., \tau)$ is also bounded on Ω and $\operatorname{grad}_x \varphi(., \tau)$ is Lipschitz on Ω with constant $L_\tau = \tau L$ for all $\tau \in [0, 1]$. Besides, $\sup_{\tau \in [0,1]} L_\tau = L < +\infty$.

We claim that $f(x) = \max_{j=1,2} f_j(x)$. Indeed, note that $f_2(x) - f_1(x) > 0$ for $x \in (0, 1)$, $f_2(x) - f_1(x) < 0$ for $x \in (1, +\infty)$ and $f_1(1) = f_2(1)$. Thus the affine function $[0, 1] \ni \tau \mapsto \varphi(x, \tau)$ satisfies

$$\max_{\tau \in [0,1]} \varphi(x, \tau) = \begin{cases} f_1(x), & x \in (0, 1), \\ f_2(x), & x \in (1, +\infty). \end{cases}$$

and the claim follows. With that characterization for f all assumptions of Theorem 4.1 are verified, with $q = 5/16$, $c = f(3/4)$ and $\delta = 2/5$, see Example in [2]. Hence, letting $x^0 \in \mathbb{R}_{++}$ and $\bar{\lambda} > 0$ such that $x^0 \in L_f(f(q))$ and $L < \mu < \lambda_k \leq \bar{\lambda}$, the proximal point method, characterized in Theorem 4.1, can be applied for solving the above nonconvex problem.

Remark 5.1. *Function $f(x) = \max_\tau \varphi(x, \tau)$, in the above example, is nonconvex (in the Euclidean sense) when restricted to any open neighborhood containing its minimizer $x^* = 1$. Therefore, the local classical proximal point method (see [15]) cannot be applied to minimize that function. Also, as f is nonconvex in the Riemannian sense, the Riemannian proximal point method (see [11]) can not be applied to minimize that function, see Example in [2] for more details.*

6 Final Remarks

We have extended the application of the proximal point method to solve nonconvex optimization problems on Hadamard manifold in the case that the objective function is given by the maximum

of a certain infinite collection of continuously differentiable functions. Convexity of the auxiliary problems is guaranteed with the choice appropriate regularization parameters in relation to the constants of Lipschitz of the field gradients of the functions which they compose the class in subject. With regards to the Theorem 4.1, in the particular case that $\varphi(., \tau)$ is convex on Ω for $\tau \in T$, convexity of the auxiliary problems is guaranteed without need of restrictive assumptions on the regularization parameters. Besides, as observed in Remark 4.2, the additional assumptions **h2** and **h3** are satisfied whenever Ω is bounded.

References

[1] Absil, P. -A., Baker, C. G., Gallivan, K. A. *Trust-region methods on Riemannian manifolds.* To appear in Foundations of Computational Mathematics. 7 (2007), no.(3), 303-330.

[2] Bento, G. C., Ferreira, O. P., Oliveira, P. R. *Proximal point methods for a Special Class of Nonconvex Functions on Hadamard Manifolds.* arXiv:0809.2594v5 [math.OC], 37 pages (2009).

[3] Attouch, H., Bolte, J., Redont, P., Teboulle, M. *Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization.* Optimization. 53 (2004), no. 5-6, 435-454.

[4] Azagra, D., Ferrera, J. López-Mesas, M. *Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds.* Journal of Functional Analysis. 220 (2005), 304-361.

[5] Clarke, F. H. *Optimization and Nonsmooth Analysis.* Classics in applied mathematics. 5, SIAM, New York, (1983).

[6] do Carmo, M. P. *Riemannian Geometry.* Boston, Birkhauser, (1992).

[7] da Cruz Neto, J. X., Ferreira, O. P., and Lucâmbio Pérez, L. R. *Contribution to the study of monotone vector fields.* Acta Mathematica Hungarica. 94 (2002), no. 4, 307-320.

[8] da Cruz Neto, J. X., Ferreira, O. P., Lucâmbio Pérez, L. R. *Monotone point-to-set vector fields.* Dedicated to Professor Constantin Udriste. Balkan J. Geom. Appl. 5 (2000), no.1, 69-79.

[9] da Cruz Neto, J. X., Ferreira, O. P., Lucâmbio Pérez, L. R., Németh, S. Z. *Convex-and Monotone-Transformable Mathematical Programming Problems and a Proximal-Like Point Method.* Journal of Global Optimization. 35 (2006), 53-69.

[10] Ferreira, O. P., Oliveira, P. R. *Subgradient algorithm on Riemannian manifolds.* Journal of Optimization Theory and Applications. 97 (1998), no.1, 93-104.

[11] Ferreira, O. P., Oliveira, P. R. *Proximal point algorithm on Riemannian manifolds.* Optimization. 51 (2000), no. 2, 257-270.

[12] R.E. Greene, H. Wu. On the subharmonicity and plurisubharmonicity of geodesically convex functions, *Ind. Univ. Math. J.*, **22** (1973), 641-654.

[13] Hare, W., Sagastizábal, C. *Computing proximal points of nonconvex functions.* Math. Program., Ser. B (2009) no. 116, 221-258.

[14] Hiriart-Urruty, J.-B, Lemaréchal, C. *Convex analysis and minimization algorithms I and II*, Springer-Verlag, (1993).

[15] Kaplan, A., Tichatschke, R. *Proximal point methods and nonconvex optimization.* J. Global Optim. 13 (1998), no. 4, 389-406.

[16] Ledyaev, Yu. S., Zhu, Qiji J. *Nonsmooth analysis on smooth manifolds.* Trans. Amer. Math. Soc. 359 (2007), no. 8, 3687-3732 (electronic).

[17] Martinet, B. (1970) *Régularisation, d'inéquations variationelles par approximations successives.* (French) Rev. Française Informat. Recherche Opérationnelle 4 (1970), Ser. R-3, 154-158.

[18] Motreanu, D., Pavel, N. H. *Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds.* J. Math. Anal. Appl. 88 (1982), no. 1, 116-132.

[19] Papa Quiroz, E. A., O. P. and Oliveira, P. R. *Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds.* To appear in Journal of Convex Analysis 16 (2009).

[20] Rapcsák, T. *Smooth nonlinear optimization in R^n.* Kluwer Academic Publishers, Dordrecht, (1997).

[21] Rockafellar, R. T. *Monotone operators and the proximal point algorithm.* SIAM J. Control. Optim. 14 (1976) 877-898.

[22] Sakai, T. *Riemannian geometry.* Translations of mathematical monographs, 149, Amer. Math. Soc., Providence, R.I. (1996).

[23] Smith, S. T. *Optimization techniques on Riemannian Manifolds.* Fields Institute Communications, Amer. Math. Soc., Providence, R.I. 3 (1994), 113-146.

[24] Spingarn, Jonathan E. *Submonotone mappings and the proximal point algorithm.* Numer. Funct. Anal. Optim. 4 (1981/82), no. 2, 123-150.

[25] Udriste, C. *Convex functions and optimization methods on Riemannian manifolds.* Mathematics and its Applications. 297, Kluwer Academic Publishers (1994).

[26] Thämelt, W. *Directional derivatives and generalized gradients on manifolds.* Optimization 25 (1992), no. 2-3, 97-115.